

## CLAIMS:

1. An ensemble of  $k$  different probing units, for determining, by hybridization,  $n$  different target oligonucleotides in an assayed sample; each of said probing units  
5 comprises one or more probe oligonucleotides with one or more probing nucleotide sequences and each of said target oligonucleotides comprising one or more target nucleotide sequences, with the probing nucleotide sequences being capable of hybridizing to target nucleotide sequences, characterized in that the probing nucleotide sequences of at least one probing unit can hybridize to target nucleotide  
10 sequences in at least two different target oligonucleotides.
2. An ensemble according to Claim 1, characterized in that at least one of the probe oligonucleotides has a probing sequence which is complementary to target sequences in at least two target oligonucleotides.
3. An ensemble according to Claim 1 or 2, characterized in that a plurality of  
15 said probing nucleotide sequences can each hybridize to two or more target nucleotide sequences in different target oligonucleotides.
4. An ensemble according to Claim 3, characterized in that a plurality of said probe oligonucleotides have probing sequences, which are complementary to two or more target nucleotide sequences in different target oligonucleotides.
- 20 5. An ensemble according to any one of Claims 1-4, comprising at least two probing units consisting of probe oligonucleotides with probing sequences, which can all hybridize to a target sequence of a single target oligonucleotide.
6. An ensemble according to Claim 5, characterized in that the probing units define groups, all oligonucleotide probes of the same group can hybridize to a  
25 target sequence in one target oligonucleotide, different groups sharing probe oligonucleotides between them, the number of probe oligonucleotides being shared between two groups being less than the number of probe oligonucleotides of at least one of the two groups.

WO 01/11079

PCT/IL00/00486

- 36 -

7. An ensemble according to Claim 6, characterized in that all oligonucleotide probes of the same group have each a probe sequence that is complementary to at least a portion of the target sequence of one target oligonucleotide.

SUB 85 } 8. An ensemble according to any one of Claims 1-7, wherein  $k$  is less than about  $10 \times n$ .

9. An ensemble according to Claim 8, wherein  $k$  is less than about  $4 \times n$ .

10. An ensemble according to Claim 9, wherein  $k$  is less than about  $2 \times n$ .

11. An ensemble according to Claim 10, wherein  $k$  is essentially equal to or less than about  $n$ .

10 12. An ensemble according to Claim 11, characterized in that out of said target oligonucleotides only a small fraction is expected to be expressed in each assayed sample, such that a vector  $e$ , the coordinates of which define expression of the different targets in the assayed sample, is a scarce vector..

SUB 86 } 15 13. A device comprising a substrate carrying an ensemble of target entities according to any one of Claims 1-12, with each of the probing units being at a defined location on the substrate.

14. A device according to Claim 13, being an oligonucleotide chip with each of said probing unit being located at a defined coordinate on the chip.

15. A method for designing a system for determining  $n$  target oligonucleotides,  
20  $S_1, S_2, \dots, S_n$ , in a sample, comprising:

(a) selecting or designing an ensemble of  $k$  probing units,  $P_1, P_2, \dots, P_k$ , each probing unit consisting of one or more probe oligonucleotide species having in combination one or more probing nucleotide sequences, the one or more probing nucleotide sequences of at least  
25 one of the probing units can hybridize to target nucleotide sequences in at least two different target oligonucleotides;

(b) arranging the ensemble of said probing units in a manner allowing exposure to the sample under conditions permitting hybridization between corresponding target oligonucleotide and probe  
30 oligonucleotide sequences and allowing determination of an

hybridization event and the extent of hybridization for each of the probe oligonucleotides;

- (c) devising  $T$  being an  $k \times n$  mathematical matrix consisting of components  $t_{ij}$ , in which matrix each  $t_{ij}$  denotes the affinity of hybridization of a target oligonucleotide  $S_i$  to probe oligonucleotides of probing unit  $P_j$ , under defined assay conditions (namely conditions to be eventually applied in the assay – type of medium, its content, temperature, etc.); and
- (d) designating the  $T$  matrix as being associated with said ensemble to permit its use in determining expression of each of said target oligonucleotides.

16. A method according to Claim 15, characterized in that the  $T$  matrix is determined empirically by exposing each probe oligonucleotide to the target oligonucleotide, under normalized conditions and determining the degree of hybridization of the target oligonucleotides to the probe oligonucleotide.

17. A method according to Claim 15, characterized in that the  $T$  matrix is determined by theoretical considerations based on the expected hybridization affinity of the target oligonucleotides to each of the probe oligonucleotides.

18. A method according to any one of Claims 15-17, wherein the ensemble of probing units is fixed on a solid substrate at a known coordinate on the substrate.

19. A method according to any one of Claims 15-18, characterized in that the probing units are selected using an optimization model in a computer simulation.

20. A method according to Claim 19 or 20, characterized in that the level of expression of each of the target oligonucleotides in an assayed sample can be calculated by applying the following vectorial equation (1):

$$c = Te \quad (1)$$

in which

SUB B7

WO 01/11079

PCT/IL00/00486

- 38 -

c is a k-dimensional vector of values  $c_1, c_2, \dots, c_k$ , representing the level of hybridization of target oligonucleotides to each of probing units,  $P_1, P_2, \dots, P_k$ , respectively.

SUB 87  
cont

e is an n-dimensional vector of values  $(e_1, e_2, \dots, e_j, \dots, e_n)$ , representing the level of expression of each of the target oligonucleotides  $S_1, S_2, \dots, S_n$ , respectively, and

T is an  $n \times k$  matrix of values  $t_{(i,j)}$ , each  $t_{(i,j)}$  being the expected level of hybridization of target oligonucleotide  $S_j$  with probing units  $P_i$ .

21. A method according to Claim 20, wherein the vector e is a sparse vector.

10 22. A method according to Claim 20 or 21, wherein the matrix T is a binary matrix.

SUB 88

23. A method according to Claim 20 or 21, wherein the matrix T is a non-binary matrix.

15 24. A method according to any one of Claims 15-23, wherein said ensemble is that according to any one of Claims 1-14.

25. A method for determining relative abundance of n target oligonucleotides  $S_1, S_2, \dots, S_n$  in an assayed sample comprising:

20 (a) providing an ensemble of k probing units,  $P_1, P_2, \dots, P_k$ , each probing unit consisting of one or more probe oligonucleotide species having in combination one or more probing nucleotide sequences, the probing units being selected such that at least one has aof the probing nucleotide sequences of at least one probing unit can hybridize to target nucleotide sequences in at least two different target oligonucleotides;

25 (b) exposing said ensemble to the assayed sample under hybridization-permissive conditions and measuring level of hybridization of target oligonucleotides from the assayed sample to each of the probing units;

30 (c) in a processor, devising a k-dimensional vector  $c = (c_1, \dots, c_k)$ , consisting of k coordinates  $c_j$ , with j being an integer from 1 to k,

each of coordinates  $c_j$  being either (i) a representation of the level of target oligonucleotides hybridized to probing unit  $P_j$ , or (ii) a representation of the difference between said level and a level measured in an identical ensemble exposed to a control sample in the same manner to that defined in step (b) (in the latter case the vector  $c$  is in fact a product of subtraction of two vectors consisting each of results obtained from a different sample);

- (d) in the processor, calculating an  $n$ -dimensional vector  $e$ , consisting of  $n$  coordinates  $e_i$ , each of coordinates  $e_i$  being an indication of the level of target  $S_i$  in the sample, by solving the following vector equation (1):

$$c = Te \quad (1)$$

in which  $T$  is a  $k \times n$  mathematical matrix consisting of components  $t_{ij}$ , in which matrix each  $t_{ij}$  denotes the affinity of hybridization of a target oligonucleotide  $S_i$  to probe oligonucleotide  $P_j$  under the assay conditions.

26. A method according to Claim 25, comprising the following additional step:

- (e) subtracting vector  $e$  from a vector  $e_c$ , vector  $e_c$  being obtained in the same manner to vector  $e$  but with a control sample.

27. A method according to Claim 25 or 26, characterized in that the vector  $e$  is a sparse vector.

28. A method according to any one of Claims 25-27, wherein the ensemble comprises also reference probing units and level of hybridization of target oligonucleotides to each probing units is compared to the level of hybridization of the target oligonucleotides to the reference probing units.

29. A method according to any one of Claims 25-28, wherein the probing units are immobilized on a substrate, each at a defined coordinate on the substrate.

30. A method according to any one of Claims 25-29, wherein the measured level of target oligonucleotides hybridized to the probing units is compared to the measured level obtained with a control sample.

31. A system for determining relative abundance of  $n$  target oligonucleotides  $S_1, S_2, \dots, S_n$ , in an assayed sample, comprising:

- (i) an ensemble of  $k$  probing units,  $P_1, P_2, \dots, P_k$ , each probing unit consisting of one or more probe oligonucleotide species having in combination one or more probing nucleotide sequences, the probing units being selected such that at least one of the probing nucleotide sequences of at least one probing unit can hybridize to target nucleotide sequences in at least two different target oligonucleotides;
- (ii) detector for detecting a quantity indicating hybridization of a target oligonucleotide to a probing unit;
- (iii) a processor coupled to said detector for constructing, based on the detected quantity, a  $k$ -dimensional vector  $c = (c_1, \dots, c_k)$ , consisting of  $k$  coordinates  $c_j$ , with  $j$  being an integer from 1 to  $k$ , each of coordinates  $c_j$  being either (i) a representation of the level of target oligonucleotides hybridized to probing unit  $P_j$ , or (ii) a representation of the difference between said level and a level measured in an identical ensemble exposed to a control sample in the same manner to that defined in step (b); and for calculating an  $n$ -dimensional vector  $e$ , consisting of  $n$  coordinates  $e_i$ , each of coordinates  $e_i$  being an indication of the level of target  $S_i$  in the sample, by solving the following vector equation (1):

$$c = Te \quad (1)$$

in which  $T$  is a  $k \times n$  mathematical matrix consisting of components  $t_{ij}$ , in which matrix each  $t_{ij}$  denotes the affinity of hybridization of a target oligonucleotide  $S_i$  to probe oligonucleotide  $P_j$  under the assay conditions.

32. A system according to Claim 31, wherein said ensemble is defined in any one of Claims 2-14.

33. For use in an assay for determining relative abundance of  $n$  target oligonucleotides  $S_1, S_2, \dots, S_n$ , in an assayed sample, a combination comprising:

- 5 (i) an ensemble of  $k$  probing units,  $P_1, P_2, \dots, P_k$ , each probing unit consisting of one or more probe oligonucleotide species having in combination one or more probing nucleotide sequences, the probing units being selected such that at least one of the probing nucleotide sequences of at least one probing unit can hybridize to a target nucleotide sequences in at least two different target oligonucleotides;
- 10 (ii) a computer readable medium carrying data for inputting to a processor, which processor, based on an inputted data constructs a vector  $c = (c_1, \dots, c_k)$ , consisting of  $k$  coordinates  $c_j$ , with  $j$  being an integer from 1 to  $k$ , each of coordinates  $c_j$  being either (i) a value representing the level of target oligonucleotides hybridized to probing unit  $P_j$ , or (ii) a value representing the difference between said level and a level measured in an identical ensemble exposed to a control sample in the same manner to that defined in step (b); calculates an
- 15  $n$ -dimensional vector  $e$ , consisting of  $n$  coordinates  $e_i$ , each of coordinates  $e_i$  being an indication of the level of target  $S_i$  in the sample, by solving the following vector equation (1):
- $$c = Te \quad (1)$$
- 20 in which  $T$  is a  $k \times n$  mathematical matrix consisting of components  $t_{ij}$ , in which matrix each  $t_{ij}$  denotes the affinity of hybridization of a target oligonucleotide  $S_i$  to probe oligonucleotide  $P_j$  under the assay conditions; said data on said data carrier comprises said matrix  $T$  which is associated for use with said ensemble.

- SUB B11 25 34. A combination according to Claim 33, wherein said ensemble is defined by any one of Claims 2-14.